

# Assessing the reliability of particle size measurements obtained by image analysis

Paul A. Larsen,\* and James B. Rawlings\*

*Department of Chemical and Biological Engineering,  
University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706*

## Abstract

The absolute accuracy of particle size distribution measurements obtained using in situ video microscopy and image analysis depends on several factors, including the focal depth, the field of view, the solids concentration, the PSD, the particle aspect ratio, and the hydrodynamics of the imaged volume. We propose a single dimensionless number that correlates with the reliability of the inferred PSD based on the likely number of overlaps per particle. As this dimensionless group becomes large, the degree of overlap and occlusion in the images makes image analysis difficult and unreliable. This parameter can aid practitioners in finding the proper sampling conditions to obtain accurate PSD measurements.

Keywords: Crystallization, Image analysis, Imaging, Morphology, Pharmaceuticals, Process monitoring.

---

\* Authors for correspondence. (P.A.L.) Telephone: (608) 265-8607. E-mail: palarsen@wisc.edu. (J.B.R.) Telephone: (608) 263-5859. Fax: (608) 265-8794. E-mail: jbraw@bevo.che.wisc.edu.

## I. INTRODUCTION

Advanced control of crystal shape, size distribution, and polymorphic form (internal lattice structure) in suspension crystallization processes has been hindered by the limitations of available on-line sensors. High-speed, in situ video microscopy is a promising technology for measuring these critical solid-phase properties. However, automatically extracting the desired measurements from in situ images in a robust and efficient manner remains challenging. We previously developed a novel image analysis algorithm called SHARC that automatically extracts particle size information from in situ images of needle-like crystals [3]. Using images acquired during a pharmaceutical crystallization experiment, we demonstrated that SHARC's PSD measurements are consistent with measurements obtained through manual, human analysis of the images, provided the images are acquired under suitable conditions. The accuracy of the measured PSD, therefore, was established only with respect to the PSD measured by human operators. Furthermore, the "suitable conditions" were characterized only in terms of the solids concentration. The absolute accuracy of the measured PSD, however, depends on several factors in addition to the solids concentration, including the focal depth, the field of view, the PSD, the particle aspect ratio, and the hydrodynamics of the imaged volume. In this work, we characterize the effects of all these factors on the absolute accuracy of the PSD measured by SHARC. The characterization is given in terms of a single variable that lumps various process and imaging conditions.

The paper is organized as follows. Section II presents the theory on which we base our characterization of SHARC's measurement accuracy. Section III describes the simulation studies used to determine the conditions for which SHARC gives an accurate PSD measurement. Section IV presents the results of these simulation studies, and Section V summarizes our findings.

## II. THEORY

Measurement reliability depends on various factors, including the optics (camera resolution  $R$ , field of view  $A$ , depth of field,  $d$ ) and the process conditions (solids concentration (w/v)  $S_w$ , particle length  $L$ , and particle shape  $k_v$ ). Most of these factors can be lumped into a single parameter denoting the expected number of crystals in the image:

$$N_c = \frac{S_w A d}{\rho_c k_v L^3}$$

Neither the solids concentration nor the number of crystals appearing in the image are sufficient to characterize the measurement reliability, however. As an example, Figures 1 and 2 show images generated for two different monodisperse particle populations at the same solids concentration and number of particles per image, respectively. For the same solids concentration, the images of the smaller particle population are more difficult to analyze, while for the same number of particles per image, the images of the larger particle population are more difficult.

The cause of measurement failure is due primarily to particle overlap. Therefore, we must calculate the probability of particle overlap, which depends not only on the number of crystals but also on the lengths and widths of the crystals. We can calculate the probability that two crystals placed randomly in the image will be overlapping as follows. Consider an image of horizontal dimension  $a$  and vertical dimension  $b$ , as shown in Figure 3. Now consider

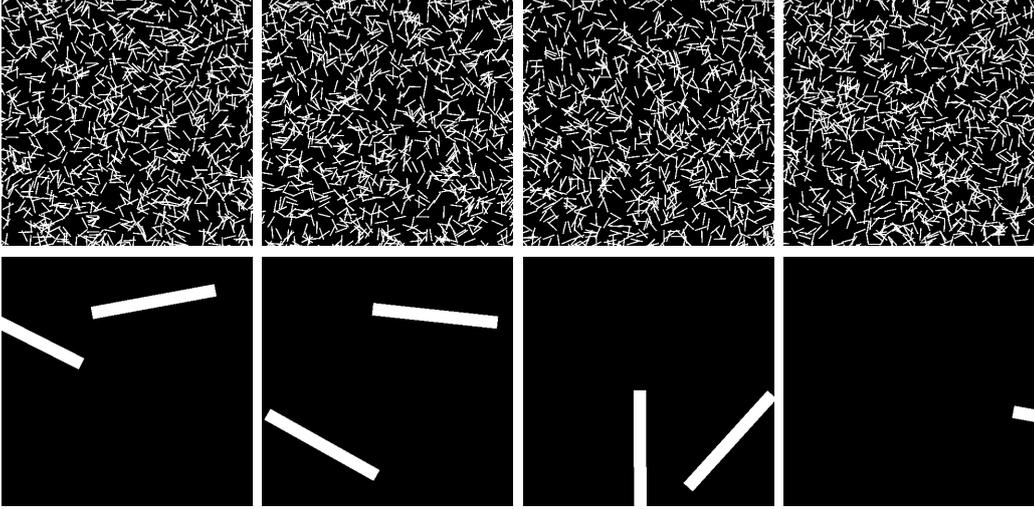


FIG. 1: Comparison of images generated for two different monodisperse particle populations at the same solids concentration. The top row shows the images for particles with length 1/20th of the image dimension and the bottom row shows images for particles with length one-half the image dimension. The aspect ratio of the particles is 10.

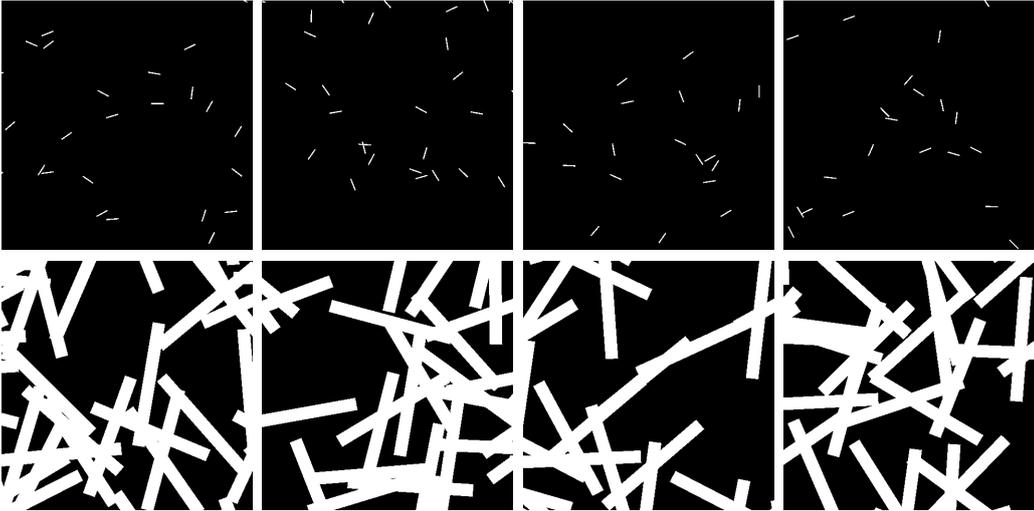


FIG. 2: Comparison of images generated for two different monodisperse particle populations at the same number of particles per image. The top row shows the images for particles with length 1/20th of the image dimension and the bottom row shows images for particles with length one-half the image dimension. The aspect ratio of the particles is 10.

a rectangular particle of length  $l$  and width  $w$  placed in the center of the image. Finally, consider a second particle of identical dimensions placed in the image with orientation  $\theta$ . If this second particle is placed anywhere inside the shaded region in Figure 3, the two particles will be overlapping. Using straightforward geometric relationships, the area  $A_{\text{ovp}}$  of this shaded region can be shown to be

$$A_{\text{ovp}} = (w(1 + \sin \theta) + l \cos \theta)(l(1 + \sin \theta) + w \cos \theta) - l^2 \cos \theta \sin \theta - w^2 \cos \theta \sin \theta \quad (1)$$

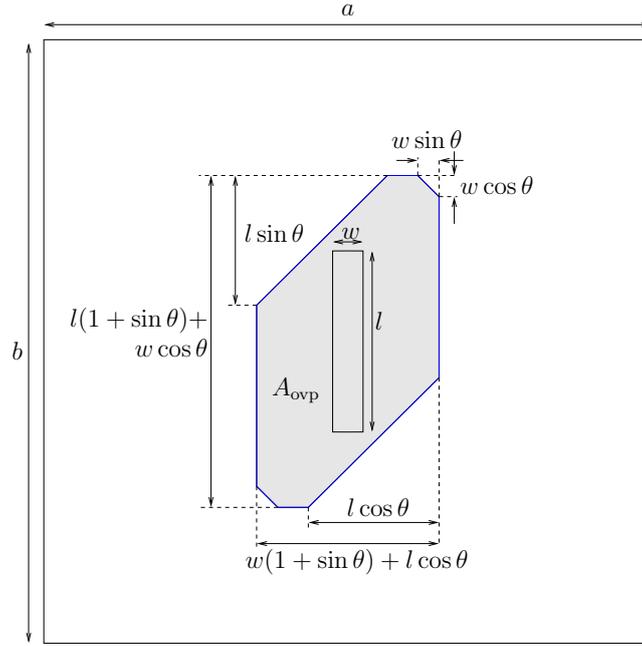


FIG. 3: Geometric representation of region in which particles of specific orientation and shape overlap.

The probability that two particles placed randomly in the image will be overlapping is

$$p_{\text{ovp}} = \frac{\int_0^{\pi/2} A_{\text{ovp}}(\theta) d\theta}{\int_0^{\pi/2} abd\theta} \quad (2)$$

This equation assumes that at least one of the particles is sufficiently far from the image boundary that Equation (1) holds. Evaluating Equation (2) gives

$$p_{\text{ovp}} = \frac{2}{\pi ab} (L^2 + w^2 + Lw(2 + \pi)) \quad (3)$$

The derivation of this equation neglects edge effects and assumes the crystals have identical size and shape and are uniformly, randomly oriented in the plane. The usefulness of this equation in assessing image difficulty, therefore, may decline as the crystal length increases (i.e. as edge effects become more significant), as the number of crystals increases, and as the dispersity in particle size increases. We expect this equation to be useful, however, for reasonable levels of particle overlap. As an indicator of the degree of difficulty of an image, we define the parameter  $D$  as

$$D = (N_c - 1)p_{\text{ovp}} \quad (4)$$

such that  $D = 0$  corresponds to an average of  $N_c = 1$  crystal per image.

The utility of the parameter  $D$  may appear limited due to its dependence on particle size and shape, which is what we're trying to measure. Hence, we cannot use  $D$  to assess the reliability of our measurement without knowing *a priori* the true value of what we're trying to measure. In practice, however, we can usually start taking measurements at low  $D$ , since batch processes progress from small crystals at low solids concentration to larger crystals at higher solids concentration. As the process progresses, we can monitor the value of  $D$  to see

when it approaches the limit at which we no longer have confidence in our size and shape measurement. From that point on, we no longer have confidence in our value of  $D$ , even though it may drop back down below the critical limit.

### III. SIMULATION STUDIES

To determine the conditions under which reliable measurements can be obtained using image analysis, we applied the SHARC algorithm described in [3] to artificial images generated at various  $D$ . SHARC is designed to identify high-aspect-ratio particles in the presence of particle overlap, and its effectiveness for real images relative to manual image analysis by human operators has been established previously [3].

The artificial images are generated using the following procedure:

1. Specify the image difficulty  $D$ .
2. Specify the image horizontal dimension  $a$  and vertical dimension  $b$ .
3. Specify the aspect ratio  $AR$  and the minimum and maximum crystal lengths,  $L_{min}$  and  $L_{max}$ , respectively.
4. Calculate the average crystal length  $\bar{L} = (L_{min} + L_{max})/2$  and average crystal width  $\bar{w} = \bar{L}/AR$ .
5. Calculate the average number of crystals in the imaging area, or  $N_c$ , using Equation (4).
6. Calculate the number density of crystals in the imaging area as  $\rho = N_c/(ab)$ .
7. Calculate an augmented image area  $A_{aug} = (a + L_{max}) \times (b + L_{max})$ . Any crystal whose centroid lies outside this augmented area has zero probability of intersecting the image boundary.
8. Calculate the average number of crystals in the augmented image area,  $N_{c,aug} = \rho A_{aug}$ .
9. Determine an integer-valued number of crystals to place in the augmented image area by sampling from a Poisson distribution with parameter  $N_{c,aug}$ .
10. Place crystal centroids randomly in image area according to uniform distributions in the horizontal and vertical dimensions.
11. Orient crystals randomly according to uniform distribution.
12. Assign crystal lengths according to chosen distribution.

SHARC was applied to the artificial images using the parameters shown in Table I.

TABLE I: Parameter values used to analyze images from crystallization experiment.

Line finder Parameters		Collinearity Thresholds		Parallelism Thresholds	
$n_{\nabla}$	5	$\epsilon_{\theta_C}$	20 degrees	$\epsilon_{\theta_P}$	5 degrees
$\epsilon_{ \nabla }$	1	$\epsilon_{EP}$	0.5	$\epsilon_Q$	0.85
$n_b$	6 buckets	$\epsilon_{PD}$	0.5	$\epsilon_{AR}$	4.5
$\epsilon_A$	20 pixels				

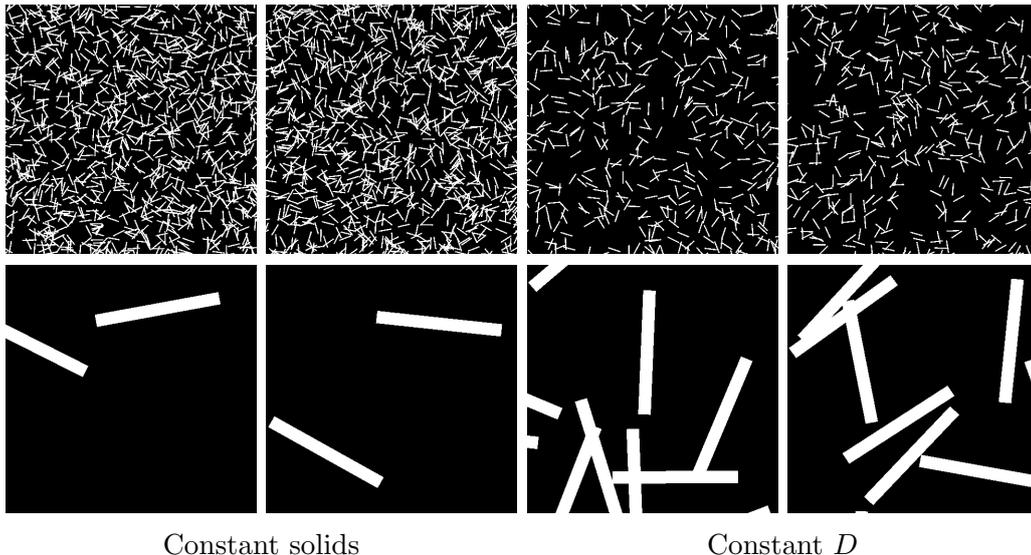


FIG. 4: Comparison of images generated for two different monodisperse particle populations at the same solids concentration and at the same  $D$ . The top row shows the images for particles with length  $1/20$ th of the image dimension and the bottom row shows images for particles with length one-half the image dimension. The aspect ratio of the particles is 10.

#### IV. RESULTS

Figure 4 compares images generated for two different monodisperse particle populations generated at the same solids concentration and at the same  $D$ . The images generated at constant  $D$  appear to be more similar in the amount of particle overlap than the images generated at constant solids concentrations.

Simulations at various  $D$  were carried out to determine the level of difficulty at which SHARC's measurement is unreliable. The simulations were carried out assuming a uniform distribution in length on  $[0.1a \ 0.99a]$  with a constant aspect ratio of 18. For each simulation, the number of measurable particles appearing within the imaging domain was approximately 10,000. Figure 5 shows four example images from each simulation.

Figure 6(a) shows the histograms obtained for each simulation assuming the image analysis is perfect, that is, assuming every particle appearing entirely within the image is identified and sized perfectly. Because larger particles are less likely to appear entirely within the image, the particle count decreases as particle size increases. This bias can be corrected by applying several different methods [2]. Figure 6(b) shows the normalized histogram obtained by applying the Miles-Lantuéjoul correction [1, 4]. As the ratio of the particle length to the

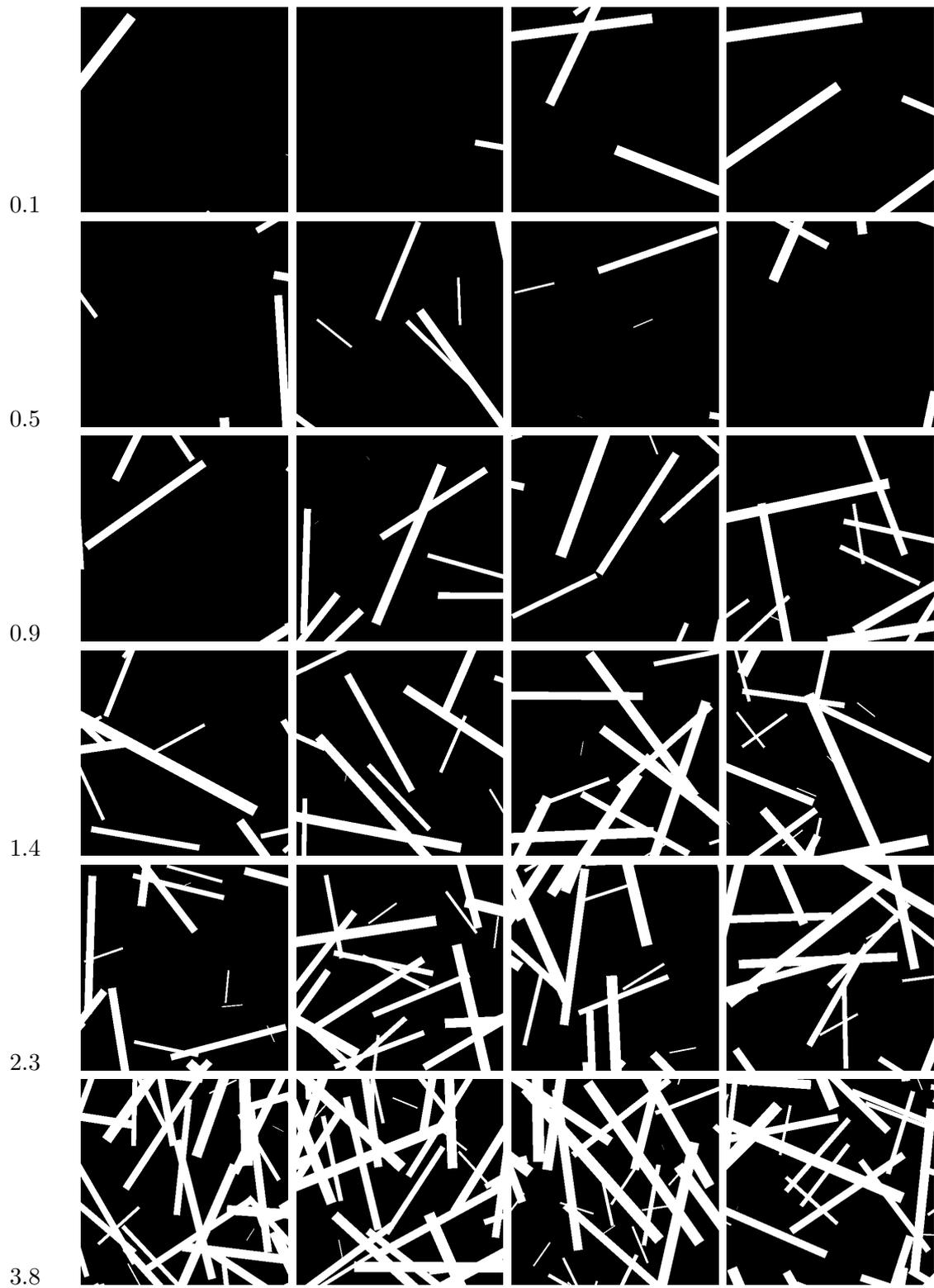


FIG. 5: Examples of synthetic images generated at various  $D$ .

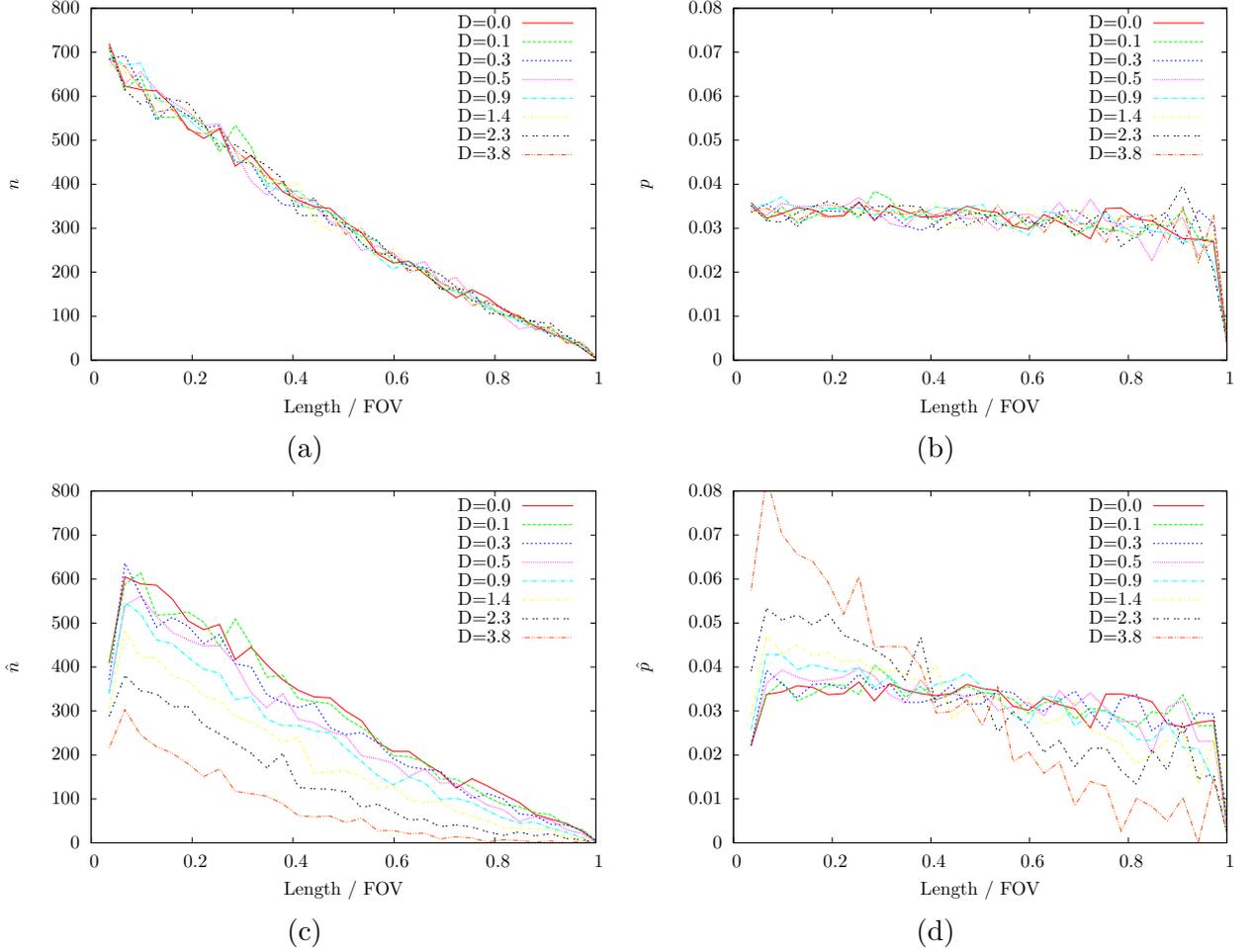


FIG. 6: Non-corrected and corrected histograms generated by analyzing artificial images at various  $D$ . (a) Non-corrected histogram obtained assuming perfect image analysis. (b) Corrected, normalized histogram assuming perfect image analysis. (c) Non-corrected histogram obtained using SHARC. (d) Corrected, normalized histogram obtained using SHARC.

field of view (FOV), or image dimension, becomes large, the accuracy of the measurement declines, even for perfect image analysis. Figures 6(c) and (d) indicate that the reliability of SHARC's measurement declines dramatically around  $D = 1.0$ .

## V. CONCLUSION

The image difficulty parameter  $D$  is useful for characterizing the difficulty associated with analyzing an image in terms of the process and imaging conditions. This parameter can be used to aid practitioners in determining the sampling conditions necessary to obtain reliable PSD measurement.

## **VI. ACKNOWLEDGMENT**

This material is based upon work supported by the National Science Foundation under Grant No. 0540147.

- 
- [1] Ch. Lantuéjoul. Computation of the histograms of the number of edges and neighbours of cells in a tessellation. In R.E. Miles and J. Serra, editors, *Geometrical Probability and Biological Structures: Buffon's 200th Anniversary*, number 23 in Lecture Notes in Biomathematics, pages 323–329, Berlin-Heidelberg-New York, 1978. Springer-Verlag.
- [2] Paul A. Larsen and James B. Rawlings. Maximum likelihood estimation of particle size distribution for high-aspect-ratio particles using in situ video imaging. In preparation.
- [3] Paul A. Larsen, James B. Rawlings, and Nicola J. Ferrier. An algorithm for analyzing noisy, in situ images of high-aspect-ratio crystals to monitor particle size distribution. *Chemical Engineering Science*, 61(16):5236–5248, 2006.
- [4] R.E. Miles. *Stochastic Geometry*, chapter 3.5, pages 228–247. John Wiley & Sons, 1974.